HHS Public Access

Author manuscript

J Cogn Neurosci. Author manuscript; available in PMC 2015 March 31.

Published in final edited form as:

J Cogn Neurosci. 2014 May ; 26(5): 1141–1153. doi:10.1162/jocn_a_00556.

Distributed and Dynamic Storage of Working Memory Stimulus Information in Extrastriate Cortex

Kartik K. Sreenivasan[1,2], **Jason Vytlacil**[1], and **Mark D'Esposito**[1]

[1] Helen Wills Neuroscience Institute University of California, Berkeley

Abstract

The predominant neurobiological model of working memory (WM) posits that stimulus information is stored via stable elevated activity within highly selective neurons. Based on this model, which we refer to as the canonical model, the storage of stimulus information is largely associated with lateral prefrontal cortex (lPFC). A growing number of studies describe results that cannot be fully explained by the canonical model, suggesting that it is in need of revision. In the present study, we directly test key elements of the canonical model. We analyzed functional MRI data collected as participants performed a task requiring WM for faces and scenes. Multivariate decoding procedures identified patterns of activity containing information about the items maintained in WM (faces, scenes, or both). While information about WM items was identified in extrastriate visual cortex (EC) and lPFC, only EC exhibited a pattern of results consistent with a sensory representation. Information in both regions persisted even in the absence of elevated activity, suggesting that elevated population activity may not represent the storage of information in WM. Additionally, we observed that WM information was distributed across EC neural populations that exhibited a broad range of selectivity for the WM items rather than restricted to highly selective EC populations. Finally, we determined that activity patterns coding for WM information were not stable, but instead varied over the course of a trial, indicating that the neural code for WM information is dynamic rather than static. Together, these findings challenge the canonical model of WM.

Introduction

Early single-unit investigations into the neural basis of working memory (WM) documented elevated firing in neurons in lateral prefrontal cortex (lPFC) when a monkey was required to store information online to link a stimulus to a subsequent response (Funahashi, Bruce, & Goldman-Rakic, 1989; Fuster, 1973; Fuster & Alexander, 1971; Kubota & Niki, 1971). This activity, termed ‘delay period activity’, has been interpreted by many (though, notably, not Fuster & Alexander, 1971 or Kubota & Niki, 1971) as representing the short-term maintenance of information about the to-be-remembered stimulus. These observations inspired a highly influential theoretical framework that has motivated several seminal findings in the study of WM and continues to shape the scope and tenor of WM research

[2] Corresponding author: kartik.sreenivasan@nyu.edu.
KKS's current affiliation: Department of Science & Mathematics New York University Abu Dhabi

(Goldman-Rakic, 1995; Wilson, Scalaidhe, & Goldman-Rakic, 1993). We refer to this framework as the canonical model of WM.

There are several key tenets of the canonical model of WM. One tenet that is the subject of recent debate is the notion that lPFC neurons store information about the sensory features of memoranda in the service of WM. This view has been bolstered by the consistent observation of delay period activity in lPFC. However, recently developed multivariate decoding methods, which rely on supervised learning algorithms to identify patterns of brain activity that represent specific types of information (Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006), offer potentially increased sensitivity relative to traditional univariate methods for localizing information content (Jimura & Poldrack, 2012). These methods have increasingly been applied to the study of how information is represented in WM (Sreenivasan, Curtis, & D'Esposito, in press). Several functional MRI (fMRI) studies utilizing decoding methods have identified patterns of visual activity that code for sensory properties of visual items during WM for those items (Christophel, Hebart, & Haynes, 2012; Ester, Serences, & Awh, 2009; Han, Berg, Oh, Samaras, & Leung, 2013; Harrison & Tong, 2009; Linden, Oosterhof, Klein, & Downing, 2011; Riggall & Postle, 2012; Serences, Ester, Vogel, & Awh, 2009; Xing, Ledgeway, McGraw, & Schluppeck, 2013). Moreover, information about maintained visual items persists in visual cortex throughout the delay period, suggesting that sensory regions participate in the storage of WM information (Harrison & Tong, 2009; Riggall & Postle, 2012). At the same time, data from single-unit studies and one recent fMRI study indicates that multivariate patterns of lPFC activity also encode information about currently maintained visual WM stimuli (S.-H. Lee, Kravitz, & Baker, 2013; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008; Rigotti et al., 2013; Stokes et al., 2013). Thus, the respective roles of these regions is unresolved. A critical step in resolving the contributions of these regions to WM involves dissociating representations that code for sensory features from those that code for non-sensory features of WM items in order to clarify the nature of the information stored in these regions.

Another tenet of the canonical model is that WM information is encoded by neural populations that are highly selective for the maintained information. In line with this view, univariate analyses of WM data have largely focused on neural populations that respond preferentially to the features of the memoranda. However, in other contexts such as the formation of sensory representations during stimulus perception, information about stimulus properties is coded for by activity in populations with a wide range of selectivity for the properties of the stimulus being represented (D. D. Cox & Savoy, 2003; Ewbank, Schluppeck, & Andrews, 2005; Haxby et al., 2001; O'Toole, Jiang, Abdi, & Haxby, 2005). It remains unclear whether WM representations similarly recruit non-selective neural populations.

Perhaps the most central tenet of the canonical model is the idea that elevated, sustained delay period activity is the neural mechanism supporting the storage of WM information. Delay period activity is consistently demonstrated in monkey electrophysiological data as well as fMRI studies in humans (e.g., Courtney, Ungerleider, Keil, & Haxby, 1997; Zarahn, Aguirre, & D'Esposito, 1999), and has become synonymous with the storage of information in WM. A corollary of this property is that WM information is coded for in a static manner

over the course of maintenance. That is, storage-related neural activity must persist in a stable form to hold WM representations in an active state. Accordingly, disruptions of delay period activity over time or due to external interference are thought to indicate a corruption of WM storage. Thus, inferences about a region's contribution to WM storage typically depend on the magnitude and temporal stability of delay period activity within that region (Artchakov et al., 2009; Jha & McCarthy, 2000; Miller, Erickson, & Desimone, 1996; Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002; Schluppeck, Curtis, Glimcher, & Heeger, 2006). The relationship between temporally stable delay period activity and WM storage is called into question by recent work that finds evidence for WM information in regions that do not exhibit delay period activity (e.g., Serences et al., 2009). While compelling, these findings do not preclude the possibility that subpopulations of voxels within their regions of interest exhibit robust delay period activity and disproportionately encode WM information. In addition, studies examining population coding of sensory features have observed that information about sensory features is maximal during temporally varying patterns of activation rather than periods of stable population activity (Mazor & Laurent, 2005).

Taken together, the evidence outlined above necessitates a reevaluation of the canonical model of WM. The goal of the present study was to critically evaluate key elements of this model. We analyzed fMRI data from 49 healthy adult participants who performed a delayed recognition task requiring WM for faces, scenes, or both faces and scenes, depending on task instructions. First, we investigated the respective roles of lPFC and visual cortex during WM by directly comparing the nature of the information encoded by these two regions. Next, we systematically examined the degree to which sensory representations of WM stimuli are dependent on activity in neural populations that are highly selective for the maintained items. Finally, we tested the relationship between information storage and stable elevated delay period activity, and characterized the temporal properties of WM information storage.

Methods

Participants

Data from 49 healthy adult participants, aged 18-32 (mean = 22.6 years; 20 female), was included in this analysis. All participants were right-handed with normal or corrected-to-normal vision and were not taking any medications with psychoactive, cardiovascular, or homeostatic effects. Written informed consent was obtained from all participants according to procedures approved by the University of California, Berkeley Committee for Protection of Human Subjects. Analyses of portions of this dataset have previously been published elsewhere (J. R. Cohen, Sreenivasan, & D'Esposito, 2012; Gazzaley, Cooney, McEvoy, Knight, & D'Esposito, 2005).

Behavioral Task

A sample trial of the WM task is depicted in Figure 1a. Participants viewed four sequentially presented sample images (two faces and two scenes in randomized order). Each sample image was presented for 800 ms with a 200 ms interstimulus interval. Participants' task

varied according to instructions presented at the beginning of each scanning run. On *Remember Faces* trials, participants were instructed to remember the two faces and ignore the two scenes; on *Remember Scenes* trials, participants were instructed to remember the two scenes and ignore the two faces; on *Remember Both* trials, participants were instructed to remember all four sample images. Participants maintained the relevant sample images in WM over a 9 s blank delay period. Following the delay period, a single probe image was presented for 1 s, and participants responded indicating whether the probe image matched one of the relevant sample images (50% probability). The probe image was always a face on *Remember Faces* trials, and was always a scene on *Remember Scenes* trials. The probe image could be either a face or scene on *Remember Both* trials. Data from a perceptual control condition that did not require WM was not included in the analyses described here. Trials were separated by a 10 s intertrial interval. Each scanning run consisted of 10 trials of a single condition. There were three runs for each condition presented over the course of the experiment.

FMRI Data Acquisition and Preprocessing

Imaging data was collected with a 4T Varian INOVA scanner equipped with a transverse electromagnetic send-and-receive radio frequency head coil. Functional data was acquired with a two-shot T2*-weighted echoplanar imaging sequence (18 slices, slice thickness 5 mm, repetition time [TR] = 2000 ms, echo time [TE] = 28 ms, matrix 64 x 64, field of view 224 mm). Slice-time correction was applied offline using sinc interpolation. Each shot of half k-space was combined with the bilinear interpolation of the two flanking shots to result in an interpolated TR of 1000 ms. In order to register functional data to brain anatomy, a T1-weighted gradient-echo multi-slice (GEMS) anatomical scan with the same slice prescription as the functional images (TR = 200 ms, TE = 5 ms, matrix 256 x 256, field of view 224 mm) and a high-resolution anatomical 3D MP-FLASH (TR = 9 ms, TE = 5 ms, matrix 256 × 256 × 128, field of view 224 × 224 × 198 mm) were additionally acquired. Functional and anatomical data were preprocessed using FSL 4.1 (FMRIB's Software Library: www.fmrib.ox.ac.uk/fsl): the MCFLIRT module was used for motion correction and BET was used to skull-strip the data. All analyses were conducted in individual participant space on unsmoothed data.

Regions of Interest

Anatomical lPFC and extrastriate visual cortex (EC) regions of interest (ROIs) are shown in Figure 1b. ROIs were defined on a standard brain (MNI152) and transformed to individual participant space using FSL's FLIRT module for linear registration. The parameters to register the GEMS anatomical image to the high-resolution MP-FLASH anatomical image (7 degrees of freedom) and the parameters to register the MP-FLASH to standard MNI152 space (12 degrees of freedom) were combined and inverted to provide the transformation from MNI space to individual participant space. LPFC (mean size = 1680 voxels; s.e.m. = 35 voxels) was defined by combining the unthresholded templates of bilateral middle frontal gyrus and bilateral inferior frontal gyrus from the Harvard-Oxford Probabilistic Brain Atlas (FSL; provided by the Harvard Center for Morphometric Analysis). The boundaries of the bilateral EC ROI (mean = 1679; s.e.m. = 36) were determined anatomically on the standard template brain and included the lingual gyrus, the parahippocampal gyrus, posterior portions

of the fusiform and inferior temporal gyri extending rostrally to the mid-fusiform gyrus to include the typical location of the fusiform face area (Kanwisher, McDermott, & Chun, 1997), and the surrounding occipital cortex.

Univariate fMRI Analysis

Although our primary analyses involved multivariate decoding methods, we used a traditional univariate general linear model (GLM) to identify canonical delay period activity. To visualize the timecourse of the BOLD data, individual trial timeseries were extracted from each ROI, z-scored, and averaged across trials, with the first TR of each trial serving as a baseline (Fig. 1a, bottom; see Fig. 1c for timecourses separated by task condition). It should be noted that z-scored timecourses are only presented for visualization purposes; all analyses of delay period activity magnitude were conducted on parameter estimates of the GLM (below). Parameter estimation for events of interest was conducted in AFNI (R. W. Cox, 1996). Our model included regressors for sample, delay, and probe events for each task condition (9 events of interest; correct trials only). Sample and probe events were modeled as 4 s and 1 s boxcar functions located at sample and probe stimulus onset, respectively. Delay events were modeled as a 1 s boxcar function located in the middle of the delay period. Regressors for each event type were created by convolving the boxcars with a canonical gamma HRF. Previous analyses have demonstrated that this method of temporally segregating regressors by at least 4 s results in sufficiently low autocorrelation between events and can therefore produce independent parameter estimates for each regressor (Zarahn et al., 1999; Zarahn, Aguirre, & D'Esposito, 1997). This approach has successfully been used to isolate sample-evoked activity from delay-related activity (J. R. Cohen et al., 2012; Jha, Fabian, & Aguirre, 2004; Pessoa et al., 2002; Yoon, Curtis, & D'Esposito, 2006). Nuisance regressors included estimated motion parameters; sample, delay, and probe events for incorrect or missed trials; and the first and second derivatives of the gamma HRF to account for differences in the latency and dispersion of the peak BOLD response.

One of our analyses examined whether delay period activity magnitude was related to decoding evidence for the storage of WM information. In order to formally investigate this relationship, we divided each anatomical ROI into tertiles based on the magnitude of delay period activity in each voxel. The magnitude of delay period activity in a given voxel was determined by the *t*-value of the delay period parameter estimate from the GLM collapsed across the three conditions, and voxels were assigned to the top, middle, or bottom delay period tertile ROI according to this value.

Another analysis investigated the degree to which WM information was encoded by category-selective voxels. This required first defining the face- and scene-selectivity of voxels within an ROI and then removing voxels from the decoding analysis according to their selectivity. Voxels were ranked according to their preference for faces or scenes by analyzing localizer data from an independent scanning run. In this run, 16-s blocks of rapidly presented face and scene stimuli were interspersed with blank 16-s blocks, and participants were instructed to indicate stimulus repetitions with a button press. Data acquisition, preprocessing, and model (GLM) parameters were as described above, except that face, scene, and baseline events were modeled as 16-s boxcar functions convolved with

the canonical HRF. Parameter estimates for the face > scene and scene > face contrasts were used to determine the degree of voxels' preference for faces or scenes. The top *v* percentile of voxels consisted of the top *v*/2 percentile of face- and scene-preferring voxels.

Decoding Analyses

All decoding analyses were carried out using the Princeton MVPA toolbox (http://www.csbmb.princeton.edu/mvpa/) and custom scripts implemented in MATLAB (The MathWorks, Inc., Natick, MA). Prior to decoding, BOLD data from each voxel was detrended by scanning run, separated into individual trial epochs, and temporally z-scored. No explicit feature selection was implemented beyond the masking of data with anatomical ROIs. We analyzed equivalent numbers of trials across task conditions for each participant. Decoding analysis was implemented using a logistic regression classifier. Training data labeled by task condition (*Remember Faces*, *Remember Scenes*, *Remember Both*) was entered into the classifier, which constructs a model that can discriminate between conditions given the multivoxel patterns of activation as an input. The classifier was then tested on unlabeled test data. Above-chance (> 33% accuracy) ability to predict the condition indicates that the multivoxel patterns of activity contained information that discriminated between conditions. Successful decoding during the blank delay period would then indicate that information about the WM items persisted despite the lack of visual input, and would be positive evidence for stored WM representations.

Most of our decoding analyses employed a leave-one-trial-out cross validation scheme: the classifier was trained on data from all but one trial and tested on the remaining trial on each cross-validation fold. This procedure was repeated until each trial in turn served as the testing trial (Pereira, Mitchell, & Botvinick, 2009). Each cross-validation fold resulted in the assignment of a weight value to each voxel in the ROI for each of the three task conditions, indicating the degree to which the activity within that voxel contributed to the classifier's output for that condition. During testing of the classifier, the vector of voxel BOLD activity was multiplied by the vector of voxels weights for each condition, resulting in a single activation value for each of our three conditions for each cross-validation fold. The testing trial was assigned a classifier guess in a winner-take-all manner. Accuracies of classifier guesses were averaged over cross-validation folds, resulting in a decoding accuracy. We set the ridge penalty (lambda value) for the logistic regression classifier to 0.01. Other penalty values yielded highly similar decoding accuracies.

In order to examine whether WM information persisted across the trial, we used a temporally resolved decoding approach. This involved creating a classifier for each of the 24 sample points (TRs) in the trial, and testing each classifier only on data from the corresponding TR in other trials. The classifier was never trained and tested on data from the same trial. Thus, each training datapoint was separated from the closest testing datapoint by 23 TRs. As our focus was on identifying storage-related neural activity, statistical analyses focused on the epoch corresponding to the delay period, which, accounting for the hemodynamic lag of ~4-6s, was determined to be TRs 11-16 of each trial. This (relatively conservative) range was chosen to minimize the influence of sample- or probe-related activity on classifier estimates; however, results were consistent across less conservative

ranges. For all statistical comparisons, the relevant measure was averaged over the 6 delay TRs. Statistical significance of decoding accuracies was assessed with a one-sample t-test, with 33% accuracy as chance-level decoding. All comparisons were two-tailed.

One of our objectives was to investigate the nature of information encoded within lPFC and EC ROIs. We reasoned that sensory representations of more similar categories would be encoded in activity patterns that were more similar; thus, for example, patterns encoding sensory representations of faces should be more similar to patterns encoding both faces and scenes than they should be to patterns encoding scenes alone. In order to examine the similarity of patterns of activity in our task conditions, we examined misclassification rates (Chen et al., 2012; Kriegeskorte, 2008) for the *Remember Faces* and *Remember Scenes* conditions. We divided trials on which the classifier had incorrectly guessed the task condition into trials on which the classifier incorrectly guessed *Remember Both* and trials on which the classifier incorrectly guessed the opposite perceptual category (i.e., when the classifier guessed *Remember Faces* for a *Remember Scenes* trial, or when it guessed *Remember Scenes* for a *Remember Faces* trial). The proportion of incorrect classifier guesses for *Remember Both* and the opposite perceptual category were combined across *Remember Faces* and *Remember Scenes* conditions. These proportions were entered into a two-way repeated-measures ANOVA with factors of ROI (lPFC vs. EC) and classifier guess (guess *Remember Both* vs. guess opposite perceptual category).

A separate classification procedure was used to examine the temporal stability of WM population coding. Unlike the previous procedure, which involved constructing a classifier for each TR that was only tested on data from the corresponding TR in other trials, this procedure involved constructing a classifier for each TR and testing each classifier on data from each TR in turn. This temporal cross-generalization procedure (Meyers et al., 2008; Stokes et al., 2013) enabled us to determine whether patterns of activity that encoded WM information at one point during the trial encoded WM information at other points in the trial as well. Temporal cross-generalization precluded the use of a leave-one-trial-out cross-validation approach, since TR 24 of trial *n*-1 and TR 1 of trial *n* would be temporally contiguous, in violation of the rule that training and testing data should be independent to avoid biasing the classifier. Instead, we divided the dataset into six groups, each of which contained data from each trial type. The classifier was trained on five groups and tested on the sixth using a leave-one-group-out cross-validation procedure, thus ensuring that training and testing datasets were independent. Lambda was set to 100 for this analysis.

Results

Decoding WM category information

BOLD data from *Remember Faces*, *Remember Scenes*, and *Remember Both* trials was entered into a logistic regression classifier, which was trained on data labeled with the relevant WM stimulus category for each trial and tested on its ability to distinguish the relevant WM stimulus category in independent unlabeled data. The logic behind this approach is that if a region represents WM stimulus information, then our classifier should be able to distinguish between task conditions at an above-chance level. We applied the decoding analysis independently to each of the 24 TRs that comprised the data acquired

within a trial in order to examine whether evidence for WM information persisted over the course of the trial. Above-chance decoding accuracy corresponding to the delay period of the trial, when no visual information was present and WM maintenance was ongoing, was taken as evidence for the storage of WM information. Our analyses were restricted to two ROIs, lPFC and EC (Fig. 1b), that have been implicated in the storage of visual WM information (Fuster, Bauer, & Jervey, 1985; Lepsien & Nobre, 2007; Pessoa et al., 2002; Petrides, 2000; Ranganath, Cohen, Dam, & D'Esposito, 2004; Sakai, Rowe, & Passingham, 2002; Zarahn et al., 1999). The decoding analysis demonstrated robust above-chance accuracy across the trial (Fig. 2a, left), and in particular during the delay phase of the trial in both EC and lPFC ROIs ($t(48) > 7.3$; ps < 0.0001; Cohen's $d > 1.05$; Fig. 2a, right), indicating that category representations were maintained in both regions.

The nature of WM information in EC and lPFC

One of our primary goals was to distinguish WM representations that were sensory in nature, as would be expected if a region participates in WM storage, from non-sensory representations such as rules, goals, or abstract representations of categories. To do so, we examined the classifier's misclassification rates, which can provide insight into the representational similarity of our categories of interest (Chen et al., 2012; Kriegeskorte, 2008). We reasoned that if a region supports a sensory representation of WM stimuli, then *Remember Faces* trials should be incorrectly classified as *Remember Both* trials more often than they should be misclassified as *Remember Scenes* trials, since the sensory representation of faces is more similar to the representation of faces and scenes than it is to scenes. Similarly, *Remember Scenes* trials should also be disproportionately misclassified as *Remember Both* trials if activity patterns encode sensory representations. This approach was motivated by previous work demonstrating that visual neurons respond based on visual similarity to their preferred feature, while lPFC neurons encode arbitrary and abstract category boundaries independent of visual similarity (Freedman, Riesenhuber, Poggio, & Miller, 2001; 2003). Thus, our prediction was that misclassification rates in EC would be consistent with a sensory representation, while misclassification rates in lPFC would not distinguish between visually similar categories. We compared the pattern of misclassification in our two ROIs during the delay period by performing a two-way ANOVA on the proportion of misclassified trials with the factor of ROI and guess type (guess *Remember Both*, and guess opposite perceptual category – i.e., face guess on scene trials and vice versa). We found a significant ROI x guess type interaction ($F(1,48) = 10.49$, $p = 0.002$; $\eta_p^2 = 0.18$; Fig. 2b): a greater proportion of *Remember Faces* and *Remember Scenes* trials were misclassified as *Remember Both* in EC ($t(48) = 3.2$; $p = 0.003$; $d = 0.45$), whereas there was no significant difference in the proportion of trials misclassified as *Remember Both* vs. the opposite perceptual category in lPFC, suggesting that EC and not lPFC stores a sensory representation of WM items.

Contribution of selective neural populations to WM information storage

To investigate whether sensory WM representations were encoded by category-selective populations within EC, we ranked EC voxels according to their category selectivity and removed increasing numbers of voxels from the decoding analysis to determine the degree to which decoding was dependent on category-selective voxels. Similar procedures have

previously been used to determine whether representations of object categories depend on selective voxels during perception (Haxby et al., 2001) and attention (Chen et al., 2012). Face- and scene-selectivity of voxels were determined in each participant in an independent scanning run (see Methods). Figure 3a shows the top 25% of selective EC voxels in two representative participants. Note that these voxels correspond well to previously described face- and scene-dedicated processing modules in EC (Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998; Gauthier et al., 2000; Kanwisher et al., 1997). After identifying these voxels, we repeated the decoding analysis as described above after removing a percentile of the most selective voxels from the analysis. The analysis was conducted removing 5%, 25%, and 50% of the most category-selective voxels from EC. Although decoding accuracy was reduced as an increasing proportion of category-selective EC voxels were removed from the analysis (Fig. 3b, left), decoding accuracy during the delay period remained significantly above chance, even when half of the voxels in EC were removed (ts(48) > 7.8, ps < 0.0001; ds > 1.1; Fig. 3b, right). From this, we concluded that while category-selective EC voxels may code for WM information, WM storage recruits distributed EC populations with a broad range of category selectivity.

Delay period activity and WM information storage

To understand the role of delay period activity in WM storage, we investigated the relationship between our decoding metrics and the magnitude of delay period activity in lPFC and EC ROIs. Individual voxels within each ROI were assigned to strata according to the magnitude of delay period activity as determined by the delay period parameter estimates of our univariate model (see Methods). We created three strata within each ROI, with the top tertile demonstrating robust delay period activity, the middle tertile showing an absence of delay period activity, and the bottom tertile demonstrating below-baseline levels of activity during the delay (Fig. 4a). If delay period activity is related to WM information storage, then the top tertile should demonstrate greater evidence for WM information storage, as evinced by higher decoding accuracy during the delay period. Decoding analyses performed separately in each tertile ROI showed that decoding accuracy was consistent across tertiles (Fig. 4b, left) and did not differ significantly during the delay in either ROI (Fs < 0.67; ps > 0.5; η_p^2 < 0.02; Fig. 4b, right).

In a complementary analysis, we examined the relationship between the magnitude of delay period activity in a voxel and the degree to which that voxel was considered informative by the classifier during the delay period. We extracted the delay period classifier weights (see Methods) for each of the three conditions (*Remember Faces*, *Remember Scenes*, *Remember Both*) from our original decoding analysis. To arrive at a single delay period weight value per voxel per condition, weights were averaged over cross-validation folds, and then over the 6 delay TRs. Both positive and negative weight values can indicate that a voxel is highly informative to the classifier; we therefore examined the correlation between the absolute magnitude of the weights and the univariate model's estimate of delay period activity in the same condition. This yielded three correlation values per ROI, which were each averaged across participants. If the magnitude of delay period activity in a voxel is an indication of the degree to which it was informative to our decoding analysis, a positive correlation should be expected. Consistent with the analysis above, the correlation coefficients were between

-0.01 and 0, indicating no relationship between a voxel's contribution to the classifier and its delay period activity magnitude. Results were qualitatively similar when using the raw weight values. Together, these analyses present a formal dissociation between the magnitude of delay period activity and WM storage.

Temporal stability of WM information storage

The above analyses dissociate the magnitude of delay period activity and WM storage. A separate but related question is whether sustained WM representations rely on stable multivoxel patterns of activity. Patterns of voxels with a wide range of activation levels could stably encode a stimulus independent of their delay period activity magnitude. Our previous decoding analyses revealed WM category information in EC that persisted throughout the trial; however, they did not distinguish whether this information was encoded via patterns of activity that were stable throughout the trial, or whether storage was carried out by patterns of activity that shifted over the course of a trial. To investigate this question, the decoding analysis was modified to train the classifier on data from one TR and test on each of the 24 TRs in turn. This procedure was repeated such that each TR served as the training TR for one iteration of testing, resulting in a 24 by 24 matrix of decoding accuracy. If information is stored in a static or stable pattern, then a classifier trained on one TR should successfully be able to decode information on nonadjacent TRs within the trial. Instead, if information is stored dynamically in temporally varying patterns of activity, then a classifier created from data from one TR should not be able to successfully decode information about the relevant stimulus category from another part of the trial (Meyers et al., 2008; Stokes et al., 2013). As our interest was in the temporal properties of sensory representations, our analysis focused on data from the EC ROI. Decoding accuracy was above chance along the diagonal of the matrix, when the classifier was trained and tested on data from the same part of the trial, but was reduced when the classifier was trained and tested on data from different TRs (Fig. 5a & b). To formally test whether patterns were stable throughout the trial, we framed our question in terms of model selection. For each training TR, our measure of interest was the difference between the decoding accuracy from the model tested on data from the same TR (the on-diagonal element of a given row of the decoding accuracy matrix) and the average decoding accuracy from the other 23 models (the average of the off-diagonal elements of the same row of the decoding accuracy matrix). If the on-diagonal element outperformed the average of the off-diagonal elements, we took this as evidence that the pattern of information on the training TR was not sustained across the trial. We then compared the proportion of participants for whom the on-diagonal model outperformed the average of the off-diagonal models against the binomial distribution B(49,0.5) for each TR. The diagonal model significantly outperformed the average off-diagonal model at all 24 TRs (Fig. 5c, all $ps < 0.005$). Critically, the use of cross-validation to evaluate our models on independent sets of data precluded the possibility that our results were the result of a single stable pattern plus noise, and allowed us to conclude that patterns containing WM information shifted over the course of the trial.

Discussion

Our results demonstrate (i) that EC retains sensory WM representations while lPFC encodes category representations that are non-sensory in nature; (ii) that WM information is stored in patterns of activity that are distributed over voxels with a broad range of selectivity; (iii) WM storage is independent of the magnitude of population delay period activity; and (iv) patterns of activity encoding WM representations vary over the course of maintenance. Along with other work describing features of WM that are incompatible with the canonical view of WM, our findings highlight the need for a reevaluation of the neural instantiation of WM. These findings also emphasize the utility of multivariate decoding analyses of fMRI data in the study of WM.

Contrasting the roles of EC and lPFC in WM

Our results show that EC retains sensory WM representations, while lPFC retains non-sensory information. These findings are consistent with growing evidence that visual WM representations are stored in visual cortex (Christophel et al., 2012; Ester, Anderson, Serences, & Awh, 2013; Harrison & Tong, 2009; Riggall & Postle, 2012; Serences et al., 2009; Silvanto & Cattaneo, 2010; Slotnick & Thakral, 2011), as well as studies highlighting the role of lPFC in forming and maintaining categorical representations and representations of important task variables (Freedman et al., 2001; 2003; Meyers et al., 2008; Rigotti et al., 2013). The key advance of the present work is that we were able to contrast the nature of the information stored in lPFC and EC within the same task, thus clarifying the respective roles of these two regions. Critically, our results provide a potential alternative explanation for previous work indicating that sensory representations are stored in lPFC; patterns of activity associated with specific stimuli in previous work may encode categorical or rule information associated with that stimulus, and not the sensory properties themselves. It is important to note that our conclusions do not rely on a comparison of decoding accuracy across regions, which could yield spurious differences arising from vascular or other differences across ROIs that might obscure informative patterns of activity. Instead, we used misclassification rates to distinguish between the nature of patterns in two ROIs that demonstrated successful decoding, allowing us to conclude that EC stores sensory information about WM items.

How do we explain discrepancies between our findings and other work that was unable to decode WM information in lPFC (Christophel et al., 2012; Riggall & Postle, 2012)? Previous fMRI studies that were unable to decode WM information from lPFC were decoding stimulus *identity* (e.g., one of several directions of motion), while our study decoded stimulus *category* while participants maintained stimulus identity in WM. While studies decoding stimulus category have the disadvantage of not being able to identify stimulus-specific patterns of activity, given lPFC's preference for category boundaries (i.e., learned abstract distinctions) over item similarity (i.e., sensory features) (Freedman et al., 2003), it is possible that the nature of our task facilitated decoding in lPFC. In line with this notion, a recent fMRI decoding study found information about maintained visual items in visual cortex and information about maintained visual categories in lPFC (S.-H. Lee et al., 2013). While the authors interpret this dissociation as a distinction between visual and verbal WM, in light of the present results, we suggest that these findings can be interpreted

as a distinction between sensory and categorical representations. Although the present work focuses on these regions in isolation, WM likely requires coordinated activity between these regions and others, including parietal cortex and basal ganglia. Further study is required to understand the individual and collective function of these regions.

Delay period activity and WM storage

Previous decoding analyses have demonstrated successful decoding of the contents of WM in the absence of delay period activity (Linden et al., 2011; Serences et al., 2009); however, these studies did not rule out the possibility that subpopulations of voxels within their regions of interest may have exhibited delay period activity and contributed disproportionately to their decoding success. One study removed all voxels with significant delay period activity and still observed information about WM items (Riggall & Postle, 2012); however, these results do not preclude the possibility that voxels with greater magnitude delay period activity may contribute more information to a classifier. Previous work also used arbitrary significance thresholds to define delay active voxels, which may obscure the contributions of just below threshold activity. By dividing our ROIs into strata based on the magnitude of delay period activity in each voxel, we were able to demonstrate a more convincing dissociation between patterns coding for WM storage and the magnitude of delay period activity. This dissociation was strengthened by our finding that delay magnitude and voxel weights were not positively correlated. In concert with evidence that delay period activity is associated with cognitive operations besides WM (Curtis & Lee, 2010; Meyer, Qi, & Constantinidis, 2007) and successful WM in the absence of delay period activity (Offen, Schluppeck, & Heeger, 2009; Serences et al., 2009), our work suggests an independence between delay period activity and WM storage. How might WM representations be sustained without relying on delay period activity? One possibility is suggested by work showing that information can be sustained over brief intervals via rapid shifts in synaptic weights (Mongillo, Barak, & Tsodyks, 2008; Sugase-Miyamoto, Liu, Wiener, Optican, & Richmond, 2008). In such a scenario, neurons that store memory traces serve as matched filters, and stimulus- or category-specific delay activity may be a function of nonspecific input into the system rather than an index of storage *per se*.

What, then, is the function of elevated sustained delay activity frequently observed during WM? While our analyses suggest that delay period activity and WM storage are not synonymous, persistent neural throughout WM maintenance is related to WM performance (J. R. Cohen et al., 2012; Pessoa et al., 2002), and is thus an important element of WM. The strong association between delay period activity and regions of PFC that carry out complex operations such as the temporal integration of behaviorally relevant goals (Fuster, 2001), suggests that one possible function of delay period activity in lPFC may be to sustain higher-order task and goal representations (Miller & Cohen, 2001). We suggest that stable delay period activity may be one of several possible neural mechanisms for retaining information in an active state.

An important consideration in evaluating these findings is the degree of accordance between data from single-unit recordings in non-human primates and multivariate analyses of human fMRI data. The former combines excellent temporal resolution with the ability to observe

spiking activity in single neurons, while the latter has relatively coarse spatial resolution, but has the advantage of broad spatial coverage to examine population codes across wide regions of cortex. Given that fMRI voxels represent the summed activity of hundreds of thousands of neurons, as well as the uncertain relationship between neuronal spiking and BOLD activity (Cardoso, Sirotin, Lima, Glushenkova, & Das, 2012; Logothetis, 2008), we cannot rule out the possibility that significant stable delay period spiking activity exists even within voxels demonstrating low levels of sustained BOLD activity. Additionally, one must consider the differences in the tasks employed in human fMRI and monkey electrophysiological studies; the former employs delay periods lasting up to 20 seconds, while the latter typically has delay periods shorter than 5 seconds. Germane to this distinction, spiking models of delay period activity in PFC typically find that item-specific delay period activity decays after several seconds (Hansel & Mato, 2013). Ultimately, methods such as monkey fMRI and electrocorticographical recordings in humans may help reconcile some of the differences between findings in humans and monkeys, providing a more complete picture of WM.

Dynamic patterns of activation in WM

While the other aspects of WM storage that we investigated are explicit elements of the canonical view, the temporal stability of WM representations is largely an implicit property of WM models. Experimental manipulations that disrupt delay period activity, such as the presentation of distracting items during the delay period, are often used to dissociate regions that participate in storage from regions that perform auxiliary roles in WM (Artchakov et al., 2009; Miller, Li, & Desimone, 1991; 1993; Yoon et al., 2006). The strong implication in this work is that WM representations must persist in a stable form across the period of maintenance. This is in contrast to evidence from psychology suggesting that WM representations undergo changes during the period of maintenance, as evidenced by different levels of susceptibility to intrusion (Oberauer, 2001), as well as evidence that stimulus features can be encoded via dynamic population codes during perception (Crowe, Averbeck, & Chafee, 2010; Mazor & Laurent, 2005).

The temporal properties of the neural correlates of WM have not been well-studied; however, extant electrophysiological evidence from rats (Baeg et al., 2003) and monkeys (Meyers et al., 2008) indicates that population coding of WM representations can involve spatially and temporally varying patterns of activity. These empirical findings are supported by theoretical work indicating that population dynamics can support the encoding of stable representations (Druckmann & Chklovskii, 2012). A recent noteworthy study used a similar temporal cross-generalization decoding approach to investigate WM representations in monkey lPFC, and found that information was encoded in time-varying patterns of activity (Stokes et al., 2013). Interestingly, the authors observed that the patterns of activity coding for WM items were more stable during the delay period relative to other parts of the trial.

In contrast to Stokes and colleagues, we found that informative patterns of activity were not stable at any point during the trial. Additionally, our finding of dynamic population coding in EC is inconsistent with previous fMRI decoding results demonstrating that WM information is contained in stable patterns of visual cortical activity (Harrison & Tong,

2009; Riggall & Postle, 2012; Serences et al., 2009). An intriguing possibility is that these discrepancies may be explained by WM load. In our study, WM load varied between 2 and 4 items, whereas most previous studies did not tax WM load to the same degree. This possibility receives tentative support from recent findings by Emrich and colleagues, who performed a decoding analysis of fMRI data during the maintenance of multiple directions of motion (Emrich, Riggall, LaRocque, & Postle, 2013). Although their analyses did not explicitly focus on temporal cross-generalization, decoding of direction of motion did not appear to generalize across the entire trial, particularly when load was high. This finding is particularly striking when compared to previous results from the same group showing robust temporal generalization with a WM load of one (Riggall & Postle, 2012). Further work will be necessary to explicitly investigate the relationship between WM load and temporal dynamics of population coding.

References

Aguirre GK, Zarahn E, D'Esposito M. An area within human ventral cortex sensitive to “building” stimuli: evidence and implications. Neuron. 1998; 21(2):373–383. [PubMed: 9728918]

Artchakov D, Tikhonravov D, Ma Y, Neuvonen T, Linnankoski I, Carlson S. Distracters impair and create working memory-related neuronal activity in the prefrontal cortex. Cerebral Cortex. 2009; 19(11):2680–2689. doi:10.1093/cercor/bhp037. [PubMed: 19329569]

Baeg EH, Kim YB, Huh K, Mook-Jung I, Kim HT, Jung MW. Dynamics of population code for working memory in the prefrontal cortex. Neuron. 2003; 40(1):177–188. [PubMed: 14527442]

Cardoso MMB, Sirotin YB, Lima B, Glushenkova E, Das A. The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. Nature Neuroscience. 2012; 15(9):1298–1306. doi:10.1038/nn.3170.

Chen AJ-W, Britton M, Turner GR, Vytlacil J, Thompson TW, D'Esposito M. Goal-directed attention alters the tuning of object-based representations in extrastriate cortex. Frontiers in Human Neuroscience. 2012; 6:187. doi:10.3389/fnhum.2012.00187. [PubMed: 22737117]

Christophel TB, Hebart MN, Haynes JD. Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. Journal of Neuroscience. 2012; 32(38):12983–12989. doi: 10.1523/JNEUROSCI.0184-12.2012. [PubMed: 22993415]

Cohen JR, Sreenivasan KK, D'Esposito M. Correspondence Between Stimulus Encoding- and Maintenance-Related Neural Processes Underlies Successful Working Memory. Cerebral Cortex. 2012 doi:10.1093/cercor/bhs339.

Courtney SM, Ungerleider LG, Keil K, Haxby JV. Transient and sustained activity in a distributed neural system for human working memory. Nature. 1997; 386(6625):608–611. doi: 10.1038/386608a0. [PubMed: 9121584]

Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage. 2003; 19(2 Pt 1):261–270. [PubMed: 12814577]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research. 1996; 29(3):162–173. [PubMed: 8812068]

Crowe DA, Averbeck BB, Chafee MV. Rapid Sequences of Population Activity Patterns Dynamically Encode Task-Critical Spatial Information in Parietal Cortex. Journal of Neuroscience. 2010; 30(35):11640–11653. doi:10.1523/JNEUROSCI.0954-10.2010. [PubMed: 20810885]

Curtis CE, Lee D. Beyond working memory: the role of persistent activity in decision making. Trends in Cognitive Sciences. 2010; 14(5):216–222. doi:10.1016/j.tics.2010.03.006. [PubMed: 20381406]

Druckmann S, Chklovskii DB. Neuronal Circuits Underlying Persistent Representations Despite Time Varying Activity. Current Biology. 2012; 22(22):2095–2103. doi:10.1016/j.cub.2012.08.058. [PubMed: 23084992]

Emrich SM, Riggall AC, LaRocque JJ, Postle BR. Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. Journal of Neuroscience. 2013; 33(15):6516–6523. doi:10.1523/JNEUROSCI.5732-12.2013. [PubMed: 23575849]

Epstein R, Kanwisher N. A cortical representation of the local visual environment. Nature. 1998; 392(6676):598–601. doi:10.1038/33402. [PubMed: 9560155]

Ester EF, Anderson DE, Serences JT, Awh E. A neural measure of precision in visual working memory. Journal of Cognitive Neuroscience. 2013; 25(5):754–761. doi:10.1162/jocn_a_00357. [PubMed: 23469889]

Ester EF, Serences JT, Awh E. Spatially global representations in human primary visual cortex during working memory maintenance. The Journal of Neuroscience. 2009; 29(48):15258–15265. doi: 10.1523/JNEUROSCI.4388-09.2009. [PubMed: 19955378]

Ewbank MP, Schluppeck D, Andrews TJ. fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. NeuroImage. 2005; 28(1):268–279. doi: 10.1016/j.neuroimage.2005.06.036. [PubMed: 16054842]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. Science (New York, NY). 2001; 291(5502):312–316. doi:10.1126/science.291.5502.312.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. A Comparison of Primate Prefrontal and Inferior Temporal Cortices during Visual Categorization. The Journal of Neuroscience. 2003; 23(12): 5235–5246. [PubMed: 12832548]

Funahashi S, Bruce CJ, Goldman-Rakic PS. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. Journal of Neurophysiology. 1989; 61(2):331–349. [PubMed: 2918358]

Fuster JM. Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. Journal of Neurophysiology. 1973; 36(1):61–78. [PubMed: 4196203]

Fuster JM. The prefrontal cortex - an update: Time is of the essence. Neuron. 2001; 30(2):319–333. [PubMed: 11394996]

Fuster JM, Alexander GE. Neuron activity related to short-term memory. Science (New York, NY). 1971; 173(3997):652–654.

Fuster JM, Bauer R, Jervey J. Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. Brain Research. 1985; 330(2):299–307. [PubMed: 3986545]

Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW. The fusiform “face area” is part of a network that processes faces at the individual level. Journal of Cognitive Neuroscience. 2000; 12(3):495–504. [PubMed: 10931774]

Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M. Top-down enhancement and suppression of the magnitude and speed of neural activity. Journal of Cognitive Neuroscience. 2005; 17(3):507–517. doi:10.1162/0898929053279522. [PubMed: 15814009]

Goldman-Rakic PS. Cellular basis of working memory. Neuron. 1995; 14(3):477–485. [PubMed: 7695894]

Han X, Berg AC, Oh H, Samaras D, Leung H-C. Multi-voxel pattern analysis of selective representation of visual working memory in ventral temporal and occipital regions. NeuroImage. 2013; 73(C):8–15. doi:10.1016/j.neuroimage.2013.01.055. [PubMed: 23380167]

Hansel D, Mato G. Short-Term Plasticity Explains Irregular Persistent Activity in Working Memory Tasks. Journal of Neuroscience. 2013; 33(1):133–149. doi:10.1523/JNEUROSCI.3455-12.2013. [PubMed: 23283328]

Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. Nature. 2009; 458(7238):632–635. doi:10.1038/nature07832. [PubMed: 19225460]

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science (New York, NY). 2001; 293(5539):2425–2430. doi:10.1126/science.1063736.

Haynes J-D, Rees G. Decoding mental states from brain activity in humans. Nature Reviews Neuroscience. 2006; 7(7):523–534. doi:10.1038/nrn1931.

Jha AP, McCarthy G. The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. Journal of Cognitive Neuroscience. 2000; 12(Supplement 2):90–105. [PubMed: 11506650]

Jha AP, Fabian SA, Aguirre GK. The role of prefrontal cortex in resolving distractor interference. Cognitive, Affective, & Behavioral Neuroscience. 2004; 4(4):517–527.

Jimura K, Poldrack RA. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia. 2012; 50(4):544–552. doi:10.1016/j.neuropsychologia.2011.11.007. [PubMed: 22100534]

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience. 1997; 17(11):4302–4311. [PubMed: 9151747]

Kriegeskorte N. Representational similarity analysis –connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience. 2008 doi:10.3389/neuro.06.004.2008.

Kubota K, Niki H. Prefrontal cortical unit activity and delayed alternation performance in monkeys. J Neurophysiol. 1971; 34(3):337–347. [PubMed: 4997822]

Lee S-H, Kravitz DJ, Baker CI. Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nature Neuroscience. 2013; 16(8):997–999. doi:10.1038/nn.3452.

Lepsien J, Nobre AC. Attentional modulation of object representations in working memory. Cerebral Cortex. 2007; 17(9):2072. [PubMed: 17099066]

Linden DEJ, Oosterhof NN, Klein C, Downing PE. Mapping brain activation and information during category-specific visual working memory. Journal of Neurophysiology. 2011 doi:10.1152/jn.00105.2011.

Logothetis NK. What we can do and what we cannot do with fMRI. Nature. 2008; 453(7197):869–878. doi:10.1038/nature06976. [PubMed: 18548064]

Mazor O, Laurent G. Transient Dynamics versus Fixed Points in Odor Representations by Locust Antennal Lobe Projection Neurons. Neuron. 2005; 48(4):661–673. doi:10.1016/j.neuron.2005.09.032. [PubMed: 16301181]

Meyer T, Qi XL, Constantinidis C. Persistent Discharges in the Prefrontal Cortex of Monkeys Naive to Working Memory Tasks. Cerebral Cortex. 2007; 17(Supplement 1):i70–i76. doi:10.1093/cercor/bhm063. [PubMed: 17726005]

Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. Journal of Neurophysiology. 2008; 100(3):1407–1419. doi:10.1152/jn.90248.2008. [PubMed: 18562555]

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 2001; 24:167–202. doi:10.1146/annurev.neuro.24.1.167.

Miller EK, Erickson C, Desimone R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. Journal of Neuroscience. 1996; 16(16):5154. [PubMed: 8756444]

Miller EK, Li L, Desimone R. A neural mechanism for working and recognition memory in inferior temporal cortex. Science (New York, NY). 1991; 254(5036):1377–1379.

Miller EK, Li L, Desimone R. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. Journal of Neuroscience. 1993; 13(4):1460. [PubMed: 8463829]

Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. Science (New York, NY). 2008; 319(5869):1543–1546. doi:10.1126/science.1150769.

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in Cognitive Sciences. 2006; 10(9):424–430. doi:10.1016/j.tics.2006.07.005. [PubMed: 16899397]

O'Toole AJ, Jiang F, Abdi H, Haxby JV. Partially distributed representations of objects and faces in ventral temporal cortex. Journal of Cognitive Neuroscience. 2005; 17(4):580–590. [PubMed: 15829079]

Oberauer K. Removing irrelevant information from working memory: a cognitive aging study with the modified Sternberg task. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2001; 27(4):948–957.

Offen S, Schluppeck D, Heeger DJ. The role of early visual cortex in visual short-term memory and visual attention. Vision Research. 2009; 49(10):1352–1362. doi:10.1016/j.visres.2007.12.022. [PubMed: 18329065]

Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage. 2009; 45(1 Suppl):S199–209. doi:10.1016/j.neuroimage.2008.11.007. [PubMed: 19070668]

Pessoa L, Gutierrez E, Bandettini P, Ungerleider LG. Neural Correlates of Visual Working Memory: fMRI Amplitude Predicts Task Performance. Neuron. 2002; 35(5):975–987. [PubMed: 12372290]

Petrides M. Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. Journal of Neuroscience. 2000; 20(19):7496. [PubMed: 11007909]

Ranganath C, Cohen M, Dam C, D'Esposito M. Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. Journal of Neuroscience. 2004; 24(16):3917. [PubMed: 15102907]

Riggall AC, Postle BR. The Relationship between Working Memory Storage and Elevated Activity as Measured with Functional Magnetic Resonance Imaging. Journal of Neuroscience. 2012; 32(38): 12990–12998. doi:10.1523/JNEUROSCI.1892-12.2012. [PubMed: 22993416]

Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, Fusi S. The importance of mixed selectivity incomplex cognitive tasks. Nature. 2013; 497(7451):585–590. doi:10.1038/nature12160. [PubMed: 23685452]

Sakai K, Rowe JB, Passingham R. Active maintenance in prefrontal area 46 creates distractor-resistant memory. Nature Neuroscience. 2002; 5(5):479–484.

Schluppeck D, Curtis CE, Glimcher PW, Heeger DJ. Sustained activity in topographic areas of human posterior parietal cortex during memory-guided saccades. Journal of Neuroscience. 2006; 26(19): 5098–5108. doi:10.1523/JNEUROSCI.5330-05.2006. [PubMed: 16687501]

Serences JT, Ester EF, Vogel EK, Awh E. Stimulus-specific delay activity in human primary visual cortex. Psychological Science. 2009; 20(2):207. [PubMed: 19170936]

Silvanto J, Cattaneo Z. Transcranial magnetic stimulation reveals the content of visual short-term memory in the visual cortex. NeuroImage. 2010; 50(4):1683–1689. doi:10.1016/j.neuroimage.2010.01.021. [PubMed: 20079448]

Slotnick SD, Thakral PP. Memory for motion and spatial location is mediated by contralateral and ipsilateral motion processing cortex. NeuroImage. 2011; 55(2):794–800. doi:10.1016/j.neuroimage.2010.11.077. [PubMed: 21134469]

Sreenivasan KK, Curtis CE, D'Esposito M. Revisiting the role of persistent neural activity in working memory. Trends in Cognitive Science. in press.

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J. Dynamic Coding for Cognitive Control in Prefrontal Cortex. Neuron. 2013; 78(2):364–375. doi:10.1016/j.neuron.2013.01.039. [PubMed: 23562541]

Sugase-Miyamoto Y, Liu Z, Wiener M, Optican L, Richmond B. Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. PLoS Computational Biology. 2008; 4(5)

Wilson FA, Scalaidhe SP, Goldman-Rakic PS. Dissociation of object and spatial processing domains in primate prefrontal cortex. Science (New York, NY). 1993; 260(5116):1955–1958.

Xing Y, Ledgeway T, McGraw PV, Schluppeck D. Decoding Working Memory of Stimulus Contrast in Early Visual Cortex. The Journal of Neuroscience. 2013; 33(25):10301–10311. doi:10.1523/JNEUROSCI.3754-12.2013. [PubMed: 23785144]

Yoon J, Curtis CE, D'Esposito M. Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. NeuroImage. 2006; 29(4): 1117–1126. [PubMed: 16226895]

Zarahn E, Aguirre GK, D'Esposito M. Temporal isolation of the neural correlates of spatial mnemonic processing with fMRI. Cognitive Brain Research. 1999; 7(3):255–268. [PubMed: 9838152]

Zarahn E, Aguirre G, D'Esposito M. A trial-based experimental design for fMRI. NeuroImage. 1997; 6(2):122–138. doi:10.1006/nimg.1997.0279. [PubMed: 9299386]

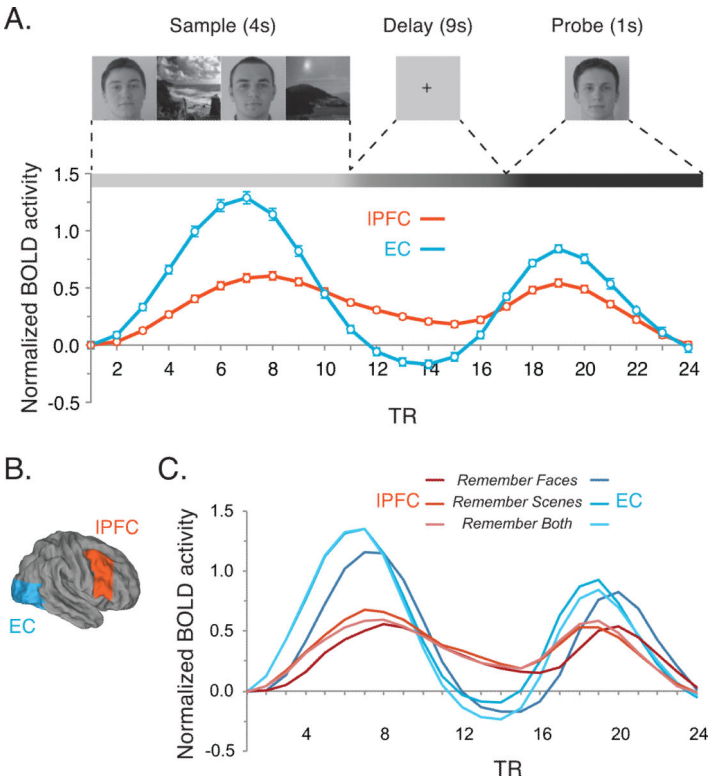

Figure 1.

Behavioral task, BOLD timecourses, and anatomical regions of interest (ROIs). **(A)** *Top:* On each trial, participants were presented with two faces and two scenes and instructed to remember the relevant sample items (faces, scenes, or both faces and scenes). The sample images were immediately followed by a blank delay period, after which participants indicated whether the probe matched one of the relevant sample items. *Bottom:* Event-related BOLD timeseries were extracted from each of the ROIs, normalized, and averaged across participants. All error bars are s.e.m. The horizontal grayscale bar indicates the phase of the trial corresponding to BOLD and decoding measures, adjusted for the convolution with the hemodynamic response function. **(B)** Analyses focused on *a priori* anatomical regions of interest: lateral prefrontal cortex (lPFC; orange), and extrastriate visual cortex (EC; blue). ROIs were bilateral; however, only the right hemisphere is shown here. **(C)** BOLD timecourses separated by task condition in lPFC and EC ROIs. Error bars have been omitted for clarity.

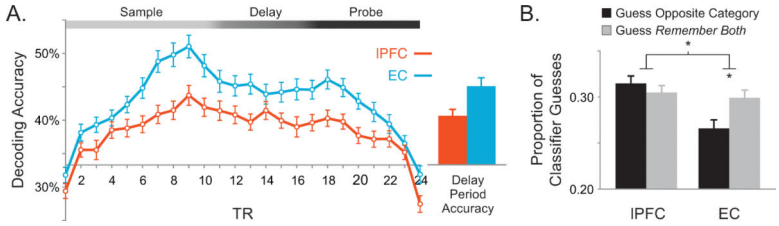

Figure 2.

Decoding WM information from lPFC and EC ROIs. **(A)** *Left:* Decoding was above chance (33% accuracy) across all three epochs of the trial in both ROIs. *Right:* To isolate information about WM items during the maintenance phase of the trial, decoding collapsed across the 6 TRs corresponding to the delay epoch of the trial. Accuracy was significantly above chance during the delay period in both ROIs, suggesting that information about the WM stimuli was maintained in both ROIs. **(B)** To distinguish between the storage of a sensory representation versus a non-sensory representation, we examined the misclassification of *Remember Faces* and *Remember Scenes* trials during the delay period. The classifier was disproportionately more likely to incorrectly guess *Remember Both* than the opposite category (i.e., guess *Remember Faces* on *Remember Scenes* trials and vice versa) in EC, consistent with a sensory representation, but not in lPFC. * indicates $p < 0.0001$.

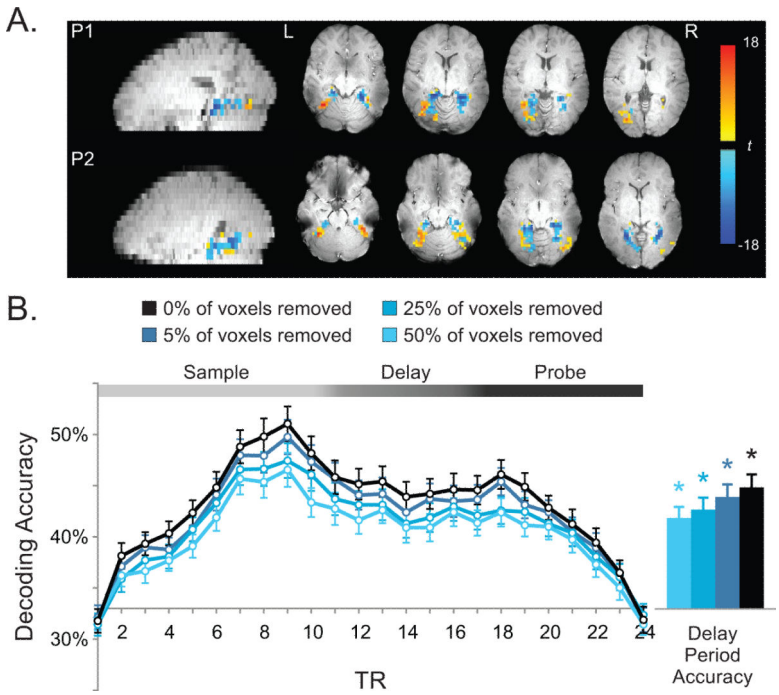

Figure 3.

Relationship between category selectivity and WM storage in EC. **(A)** An independent scanning run was used to identify category-selective voxels. The top 25% of EC voxels ranked by category selectivity are shown here for two representative participants. The contrast shown for these voxels is faces > scenes; warm colors indicate voxels selective for faces, while cool colors indicate voxels selective for scenes. **(B)** The decoding analysis was repeated as increasing percentages of voxels were removed from the ROI based on their degree of category selectivity. Accuracy remained well above chance despite the removal of up to 50% of EC voxels (decoding accuracy at each time point shown on the left; accuracy collapsed over the delay period shown on the right).

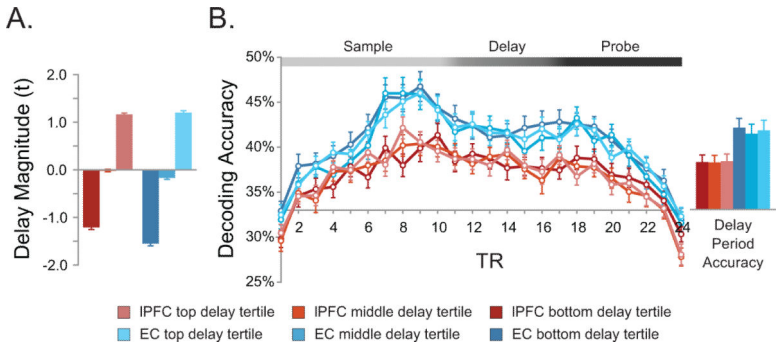

Figure 4.

Correspondence between delay period activity and WM storage. **(A)** Anatomical lPFC and EC ROIs were divided into tertiles based on the magnitude of delay period activity. The top tertile in each ROI exhibits delay period activity that is well above baseline, while the bottom tertiles show below-baseline levels of delay period activity. **(B)** Decoding was carried out separately for each tertile. Accuracy did not differ as a function of delay period magnitude in either ROI, indicating a dissociation between delay period activity magnitude and WM storage. The graph on the right depicts decoding accuracy during the delay period for both ROIs (both one-way ANOVAs with factor of tertile were non-significant).

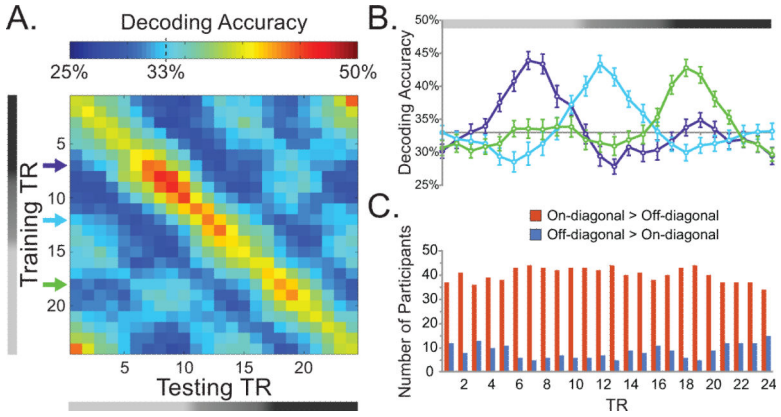

Figure 5.

Temporal dynamics of WM storage. **(A)** Temporal cross-generalization analysis involved training and testing a classifier at each of the 24 TRs that comprise a trial, resulting in a 24 × 24 matrix of decoding accuracies. Note that the diagonal of the matrix is not equivalent to the plot in Figure 2a due to slightly different decoding procedures employed in the two analyses (see Methods). **(B)** Decoding accuracy is shown for three training TRs – TR 7, TR 12, and TR 18 – indicated by the arrows in Figure 5a. **(C)** The number of participants for whom training and testing the classifier on data from the same parts of the trial (the on-diagonal elements of the matrix in A) yielded higher classification accuracy than when data from different parts of the trial (the off-diagonal elements of the matrix). At each TR, the proportion of participants with on-diagonal > off-diagonal is greater than chance (binomial test; *p*s < 0.005) indicating that patterns coding for WM representations shift across encoding, maintenance, and response.